Charles Valentine  
Homework 7  
10/23/2016

Note:  
*All code is in script hw06.R --- I have made use of cat and print statements to display information easily!*

**Problem 1**

There are 407 genes that are expressed higher in the ALL group than in the AML. This was found using the Wilcoxon two-sample test with and FDR adjustment and an alpha level of 0.05.

Using a Wlilcoxon two-sample test with an FDR adjustment we find three genes with the smallest p-values:

1. Macmarcks
2. VIL2 Villin 2 (ezrin)
3. TCF3 Transcription factor 3

This is different than the three genes with the greatest difference between means:

1. CST3 Cystatin C
2. INTERLEUKIN-8 PRECURSOR
3. Interleukin 8 (IL8) gene

This provides evidence that conclusions based only off of the point estimator mean are not as informational as when you consider the distributions of the samples.

OUTPUT

Problem 1

=========

407 genes expressed higher in the ALL group.

The three genes with the smallest p-values are:

Macmarcks

VIL2 Villin 2 (ezrin)

TCF3 Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)

The three genes with the greatest difference in means between groups are:

CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)

INTERLEUKIN-8 PRECURSOR

Interleukin 8 (IL8) gene

**Problem 2**

Using a Shapiro-Wilks normality test and FDR correction with an alpha level of 0.05, 407 genes expression levels in the AML group are not normally distributed.

OUTPUT

407 genes expressed did not pass a normality test in the AML group.

**Problem 3**

The expression levels of *HOXA9 Homeo Box A9* and *CD33* in ALL patients are significantly different with a p-value of 0.0124 found using a Wilcoxon two-sample test.

OUTPUT

HOXA9 Homeo box A9 and CD33 are significantly different (p-value: 0.0124).

**Problem 4**

To test the null hypothesis that admission and rejection rates are similar between males and females in all departments of the University we will use the Fisher exact test to calculate a p-value from the supplied contingency tables.

It is shown that the only p-value below the alpha level of 0.05 is in Department A with a p-value near 0. We can conclude that Departments B-F are not biased in accepting/rejecting students based on gender yet Department A is.

OUTPUT

Department A

P-value: 0

Department B

P-value: 0.6771

Department C

P-value: 0.3866

Department D

P-value: 0.5995

Department E

P-value: 0.3604

Department F

P-value: 0.5458

**Problem 5**

We have found using a permutation test that the variances between the ALL and AML group *CD33* expression levels are significantly different with a p-value of 0.0427 at an alpha level of 0.05.

OUTPUT

Variance in the CD333 gene expression data for ALL is smaller than in the AML group ( p-value: 0.0427 )